PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 17/30 — A1

(11) International Publication Number: WO 98/47084

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/JP98/01736

(22) International Filing Date: 16 April 1998 (16.04.98)

(30) Priority Data:

60/043,273	17 April 1997 (17.04.97)	US
08/900,214	24 July 1997 (24.07.97)	US

(81) Designated States: JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

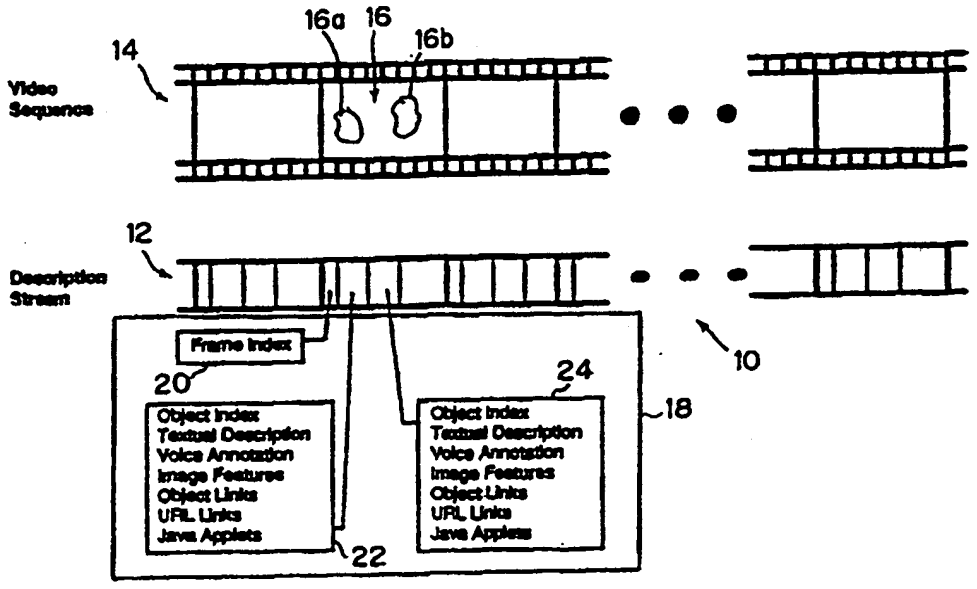
Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(71) Applicant: SHARP KABUSHIKI KAISHA [JP/JP]; 22–22, Nagaike–cho, Abeno–ku, Osaka–shi, Osaka 545–0013 (JP).

(72) Inventor: QIAN, Richard, Jungiang; Apartment 152, 501 S.E. 123 road, Vancouver, WA 98683 (US).

(74) Agents: MITANI, Hiroshi et al.; 9th floor, Salute Building, 72, Yoshida–cho, Naka–ku, Yokohama–shi, Kanagawa 231–0041 (JP).

(54) Title: A METHOD AND SYSTEM FOR OBJECT–BASED VIDEO DESCRIPTION AND LINKING



(57) Abstract

A method for object–based video description and linking is disclosed. The method constructs a companion stream for a video sequence which may be in any common format. In the companion stream, textual descriptions, voice annotation, image features, URL links, and Java applets may be recorded for certain objects in the video within each frame. The system includes a capture mechanism for generating an image, such as a video camera or computer. An encoder embeds a descriptive stream with the video and audio signals, which combined signal is transmitted by a transmitter. A receiver receives and displays the video image and the audio. The user is allowed to select whether or not the embedded descriptive stream is displayed or otherwise used.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

DESCRIPTION

A METHOD AND SYSTEM FOR OBJECT-BASED VIDEO DESCRIPTION AND LINKING

Field of the Invention

This invention relates to an object-based description and linking method and system for use in describing the contents of a video and linking such video contents with other multimedia contents.

Background of the Invention

In this information age, we daily deal with vast amount of video information when watching TV, making home video, and browsing the World Wide Web. The video which we receive or make is mostly in an "as is" state, i.e., there is no further information available about the content of the video, and the content is not linked to other related resources. Because of this, we view video in a passive manner. It is difficult for us to interact with the video contents and utilize them efficiently. From time to time, we see someone or something in the video about which we would like to find more information. Usually, we do not know where to find such information, and do not begin or continue our quest. It is also difficult for us to search for video clips which may contain certain content related to our interests.

Existing multimedia descriptive networking methods and languages comprise the known art. Examples of such methods include the descriptive techniques used in connection with digital libraries and computer languages, such as HTML and Java. The existing methods used in digital libraries suffer from shortcomings in that they are not necessarily object-based, e.g., the methods that use color histograms describe only the global color contents of a picture and do not

describe the contents of the picture; linking and networking capability is not inherent in the systems; and, the video sources must be of a specific type in order to be compatible with the primary language. Languages such as HTML and Java are difficult to use for describing and linking video contents in a video sequence, especially when it is desired to treat the video sequence at the object level.

If a video sequence were to be accompanied by a stream of descriptions and links that provided additional information about the video, and which were embedded in the video signal, we could find further information about certain objects in the video by looking up their descriptions, or visiting their related Web sites or files, by following the embedded links. Such descriptions and links may also provide useful information for content-based searching in digital libraries.

Summary of the Invention

A new method and system for object-based video description and linking is disclosed. The method constructs a companion stream for a video sequence which may be in any common format. In the companion stream, textual descriptions, voice annotation, image features, object links, URL links, and Java applets may be recorded for certain objects in the video within each frame. The method may be utilized in many applications as described below.

The system of the invention includes a mechanism for generating an encoded image. An encoder embeds a companion descriptive stream with a video signal. A video display displays the video image. The user is allowed to select whether or not the embedded descriptive stream is displayed or otherwise used.

It is an object of the invention to develop a method and system for describing and

linking video contents in any format at the video object level.

It is a further object of the invention to allow a video object to be linked to other video/audio contents, such as a Web site, a computer file, or other video objects.

These and other objects and advantages of the invention will become more fully apparent as the description which follows is read in connection with the drawings.

Brief Description of the Drawings

Fig. 1 is a block diagram of the method of the invention.

Fig. 2 is an illustration of the various types of links that may be incorporated into the invention of Fig. 1.

Fig. 3 is a block diagram of the system of the invention as used within a television broadcast scheme.

Detailed Description of the Preferred Embodiment

A new method for describing and linking objects in an image or video sequence is described. The method is intended for use with a video system having a certain digital component, such as a television or a computer. It should be appreciated that the method of the invention is able to provide additional description and links to any format of image or video. While the method and system of the invention is generally intended for use with a video sequence, such as in a television broadcast, video tape or video disc, or a series of video frames viewed on a computer, the method and system are also applicable to single images, such as might be found in an image database, and which are encoded in well-known formats, such as JPEG, MPEG, binary, etc., or any other format. As used herein, "video" includes the concept of a single "image."

Referring now to Fig. 1, the method, depicted generally at 10, builds a description

stream 12 as a companion for a video sequence 14, having a plural frames 16 therein. In each selected frame, there may be one or more objects of interest, such as object 16a and object 16b. It will be appreciated by those of skill in the art that not all of the frames in video sequence 14 must be selected for having a companion descriptive stream linked therewith.

The descriptive stream records further information about certain objects appearing in the video. The stream consists of continuous blocks 18 of data where each block corresponds to a frame 16 in the video sequence and a frame index 20 is recorded at the beginning of the block. The "object of interest" may comprise the entire video frame. Additionally, a descriptive stream may be linked to a number of frames, which frames may be sequential or non-sequential. In the case where a descriptive stream is linked with a sequential number of frames, the descriptive stream may be thought of as having a "lifespan," i.e., if the user does not take some action to reveal the descriptive stream when a linked frame is displayed, the descriptive stream "dies," and may not, in the case of a television broadcast, be revived. Of course, if the descriptive stream is part of a video tape, video disc, or computer file, the user can always return to the location of the descriptive stream and display the information. Some form of visible or audible indicia may be displayed to indicate that a descriptive stream is linked with a sequence of video frames. Descriptive stream 12 may also be linked to a single image.

The frame indexes are used to synchronize the descriptive streams with the video sequences. The block may be further divided into a number of sub-blocks 22, 24, containing what are referred to herein as descriptor/links, where each sub-block corresponds to a certain individual object of interest appearing in the frame, i.e., sub-block 22 corresponds to one object 16a in the frame and sub-block 24 corresponds to another object 16b in the same frame. There may be other

objects in the image that are not defined as objects of interest, and which, therefore, do not have a descriptive stream and sub-block associated therewith. A sub-block includes of a number of data fields including but not limited to object index, textual description, voice annotation, image features, object links, URL links, and Java applets. Additional information may include notices regarding copyright and other intellectual property rights. Some notices may be encoded and rendered invisible to standard display equipment.

The object index field is used to index an individual object within the frame. It contains the geometrical definition of the object. When a user pauses, or captures, the video at some frame, the system processes all the object index fields within that frame, locates the corresponding objects, and marks them in some manner, such as by highlighting them. The highlighted objects are those that have further information recorded. If a user "clicks" on a highlighted object, the system locates the corresponding sub-block and pop-up menu containing the available information items.

A textual description field is used to store further information about the object in plain text. This field is similar to the traditional closed caption, and its contents may be any information related to the object. The textual description can help keyword-based search for relevant video contents. A content-based video search engine may look up the textual descriptions of video sequences trying to match certain keywords. Because the textual description fields are related to individual objects, they enable truly object-based search for video contents.

A voice annotation field is used to store further information about the object using natural speech. Again, its contents may be any information related to the object.

An image features field is used to store further information about the object in terms of texture, shape, dominant color, motion model describing motion with respect to a certain reference frame, etc.. Image features may be particularly useful for content-based video/image indexing and retrieval in digital libraries.

An object links field is used to store links to other video objects in the same or other video sequence or image. Object links may be useful for video summarization and object/event tracking.

The URL links field, which is illustrated in Fig. 2, is used to store links to Web pages and/or other objects which are related to the object. For a person in the scene, such as person 26, i.e., the object of interest, the link in the sub-block 28 may be pointed to a URL 30 for the person's personal homepage 32. A symbol or icon in the scene may be linked to a Web site which contains the related background information. Companies may also want to link products 34 shown in the video, through a sub-block 36 to a URL 38 to their Web site 40 so that potential customers may learn more about their products.

A Java applet field is used to store Java code to perform more advanced functions related to the object. For example, a Java applet may be embedded to enable online ordering for a product shown in the video. Java code may also be written to implement some sophisticated similarity measures to empower advanced content-based video search in digital libraries.

In the case of digital video, the cassettes used for recording in such systems may have a solid-state memory embedded therein which serves as an additional storage location for information. The memory is referred to as memory-in-cassette (MIC). Where the video sequence is stored on a digital video cassette, the descriptive stream may be stored in the MIC, or on the

video tape. In general, the descriptive stream may be stored along with the video or image contents on the same media, i.e., a DVD disc or tape.

Figure 3 depicts the system of the invention, generally at 50, as is used in a television broadcast scheme. System 50 includes a capture mechanism, which may be a video camera, a computer capable of generating a video signal, or any other mechanism that is able to generate a video signal. A video signal is passed to an encoder 54, which also receives appropriate companion signals from the various types of links which will form the descriptive stream, which encoder generates a combined video/descriptive stream signal 58. Signal 58 is transmitted by transmitter 60, which may be a broadcast transmitter, a hard-wire system, or a combination thereof. The combined signal is received by receiver 62, which decodes the signal and generates an image for display on video display 64.

A trigger mechanism 66 is provided to cause receiver 62 to decode and display the descriptive stream. A decoder, in this embodiment, is located in receiver 62 for decoding the embedded descriptive stream. The descriptive stream may be displayed in a picture-in-picture (PIP) format on video display 64, or may be displayed on a descriptive stream display 68, which may be co-located with the trigger mechanism, which may take the form of a remote control mechanism for the receiver. Some form of indicia may be provided, either as a visible display on video display 64, or as an audible tone, to indicate that a descriptive stream is present in the video sequence.

Activating trigger mechanism 66 when a descriptive stream is present will likely result in those objects which have descriptive streams associated therewith being highlighted, or otherwise marked, to tell the user that additional information about the video object is present.

The data block information is displayed in the descriptive stream display, and the devise manipulated to allow the user to select and activate the display of additional information. The information may be displayed immediately, or may be stored for future reference. Of key importance is to allow the video display to continue uninterrupted so that others watching the display will not be compelled to remove the remote control from the possession of the user who is seeking additional information.

In the event that the system of the invention is used with a digital library, on a computer system for instance, capture mechanism 52, transmitter 60 and receiver 62 may not be required, as the video or image will have already been captured and stored in a library, which library likely resides on magnetic or optical media which is hard-wired to the video or image display. In this embodiment, a decoder to decode the descriptive stream may be located in the computer or in the display. The trigger mechanism may be combined with a mouse or other pointing device, or may be incorporated into a keyboard, either with dedicated keys, or by the assignment of a key sequence. The descriptive stream display will likely take the form of a window on the video display or monitor.

Applications

Broadcasting TV Programs

TV stations may utilize the method and system of the invention to add more functionality to their broadcasting programs. They may choose to send out descriptive streams along with their regular TV signals so that viewers may receive the programs and utilize the advanced functions described herein. The scenario for a broadcast TV station is similar to that of sending out closed caption text along with regular TV signals. Broadcasters have the flexibility of

choosing to send or not to send the descriptive streams for their programs at will. If a receiving TV set has the capability of decoding the descriptive streams, the viewer may choose to use or not use the advanced functions, just as the viewer may choose to view or not to view closed caption text. If the user chooses to use the functions, the user may read extra text about someone or something in the programs, hear extra voice annotations, or go directly to the related Web site(s), if the TV set is Web enabled, or perform some tasks, such as online ordering, by running the embedded Java applets.

For a video sequence, the descriptive stream may be obtained through a variety of mechanisms. It may be constructed manually using an interactive method. An operator may explicitly choose to index certain objects in the video and record some corresponding further information. The descriptive stream may also be constructed automatically using any video analysis tools, especially those to be developed for the Moving Pictures Experts Group Standard No. 7 (MPEG-7).

Consumer Home Video

The method and system of the invention may be utilized in making consumer video. Camcorders, VCRs and DVD recorders may be developed to allow the construction and storage of descriptive streams while recording and editing. Those devices may provide user interface programs to allow a user to manually locate certain objects in their video, index them, and recording any corresponding information into the descriptive streams. For example, a user may locate an object within a frame by specifying a rectangular region which contains the object. The user may then choose to enter some text into the textual description field, record some speech into the voice annotation field, and key in some Web page address into the URL links

field. The user may choose to allow the programming of the device to propagate those descriptions to the surrounding frames. This may be done by tracking the objects in the nearby frames. The recorded descriptions for certain objects may also be used as their visual tags.

If a descriptive stream is recorded along with a video sequence as described above, that video can then be viewed later and support all the functions as described above.

Digital Video/Image Databases

As previously noted, the method and system of the invention may also be used in digital libraries. The method may be applied to video sequences or images originally stored in any common format including RGB, D1, MPEG, MPEG-2, MPEG-4, etc. If a video sequence is stored in MPEG-4, the location information of the objects in the video may be extracted automatically. This eases the burden of manually locating them. Further information may then be added to each extracted object within a frame and propagated into other sequential or non-sequential frames, if so selected. When a sequence or image is stored in a non-object-based format, the mechanism described herein may be used to construct descriptive streams. This enables a video sequence or image stored in one format to be viewed and manipulated in a different format, and to have the description and linking features of the invention to be applied thereto.

The descriptive streams facilitate content-based video/image indexing and retrieval. A search engine may find relevant video contents at the object level, by matching relevant keywords against the text stored in the textual description fields in the descriptive streams. The search engine may also choose to analyze the voice annotations, match the image features, and/or look up the linked Web pages for additional information. The embedded Java applets may

implement more sophisticated similarity measures to further enhance content-based video/image indexing and retrieval.

Thus, a method and system for object-based video description and linking has been disclosed. It will be appreciated that variations and modifications thereof may be made within the scope of the invention as defined in the appended claims.

CLAIMS

1. A method of object-based description and linking of objects within an image, comprising:

generating a descriptive stream, including a data block, for the image;

identifying at least one object of interest in the image;

inserting description/links into the data block for an object of interest; and

recording a frame index at the beginning of each data block for synchronizing the description/links with the image.

2. The method of claim 1 wherein said inserting of description/links includes inserting description/links taken from the group of description/links consisting of object indexes, textual descriptions, voice annotation, image features, object links, URL links and Java applets.

3. The method of claim 1 wherein said identifying at least one object of interest includes identifying the entire image as an object of interest.

4. The method of claim 1 wherein the image is a portion of a sequence of images comprising a video sequence of video frames, and wherein said generating a descriptive stream includes generating a descriptive stream for plural video frames in said video sequence.

5. The method of claim 4 wherein the video frames are in sequential order in said video sequence.

6. The method of claim 4 wherein the video frames are in non-sequential order in said video sequence.

7. A method of object-based description and linking of objects within a video sequence, wherein the video sequence includes plural video frames, comprising:

generating a descriptive stream, including a data block corresponding to a select video frame in the video sequence;

identifying at least one object of interest in a video frame;

inserting description/links into the data block for an object of interest; and

recording a frame index at the beginning of each data block for synchronizing the description/links with the video sequence.

8. The method of claim 7 wherein said inserting of description/links includes inserting description/links taken from the group of description/links consisting of object indexes, textual descriptions, voice annotation, image features, object links, URL links and Java applets.

9. The method of claim 7 wherein said identifying at least one object of interest includes identifying the entire video frame as an object of interest.

10. The method of claim 7 wherein said generating a descriptive stream includes generating a descriptive stream for plural video frames in a video sequence.

11. The method of claim 10 wherein the video frames are in sequential order in a video sequence.

12. The method of claim 10 wherein the video frames are in non-sequential order in a video sequence.

13. A system for object-based video description and linking of objects to an image, wherein the image is represented by an electrical signal, comprising:

an encoder for embedding a descriptive stream with the electrical signal;

a display mechanism for displaying the image;

a decoder for decoding the embedded descriptive stream; and

a trigger mechanism for instructing said decoder to decode and display said descriptive stream in a descriptive stream display at the request of a user, and for selecting, at the request of a user, a particular portion of the descriptive stream with which to work.

14. The system of claim 13 which further includes a capture mechanism for generating the image as a sequence of video frames, and for converting said image into a video signal.

15. The system of claim 14 which further includes a transmitter for transmitting said video signal and said embedded descriptive stream; and a receiver constructed and arranged for receiving said video signal and said embedded descriptive stream and for displaying a video image.

16. The system of claim 14 wherein said capture mechanism is taken from the group consisting of video cameras and computers.

17. The system of claim 13 wherein said trigger mechanism is located in a remote-control device.

18. The system of claim 13 wherein said descriptive stream display is located in a remote-control device.

FIG.1

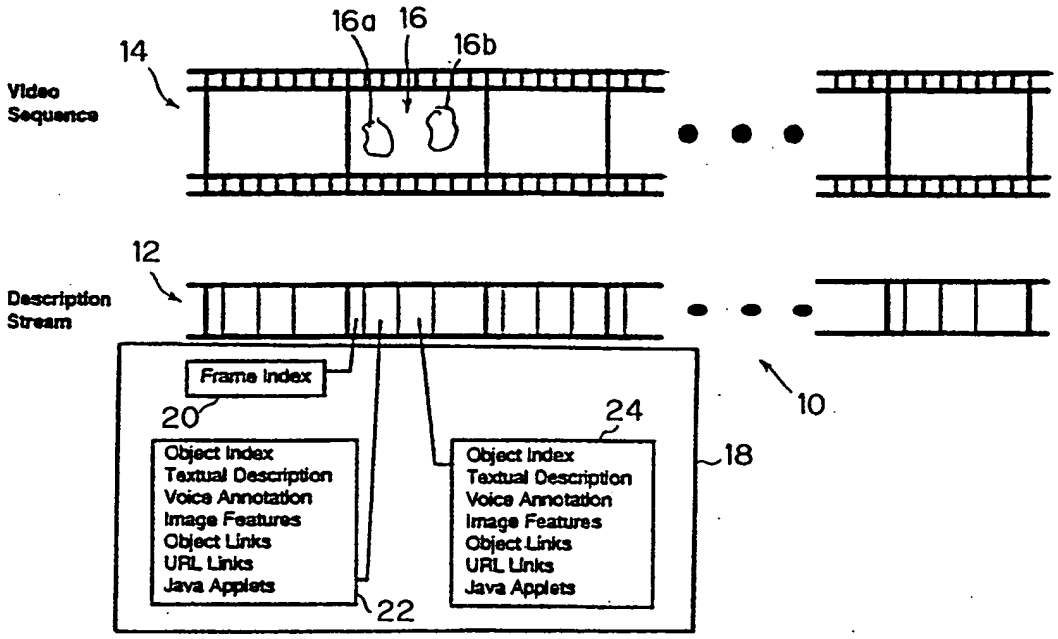

16a
16
16b
14
Video Sequence
12
Description Stream
Frame Index
20
24
Object Index
Textual Description
Voice Annotation
Image Features
Object Links
URL Links
Java Applets
22
18
10

FIG.2

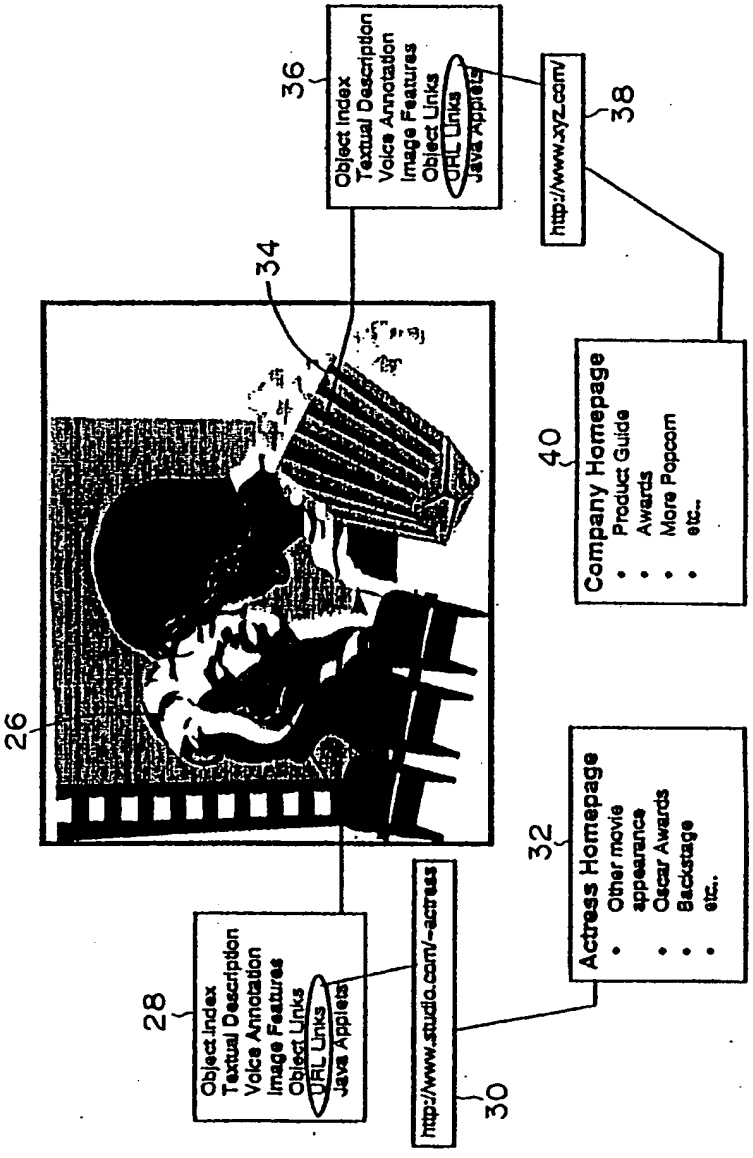
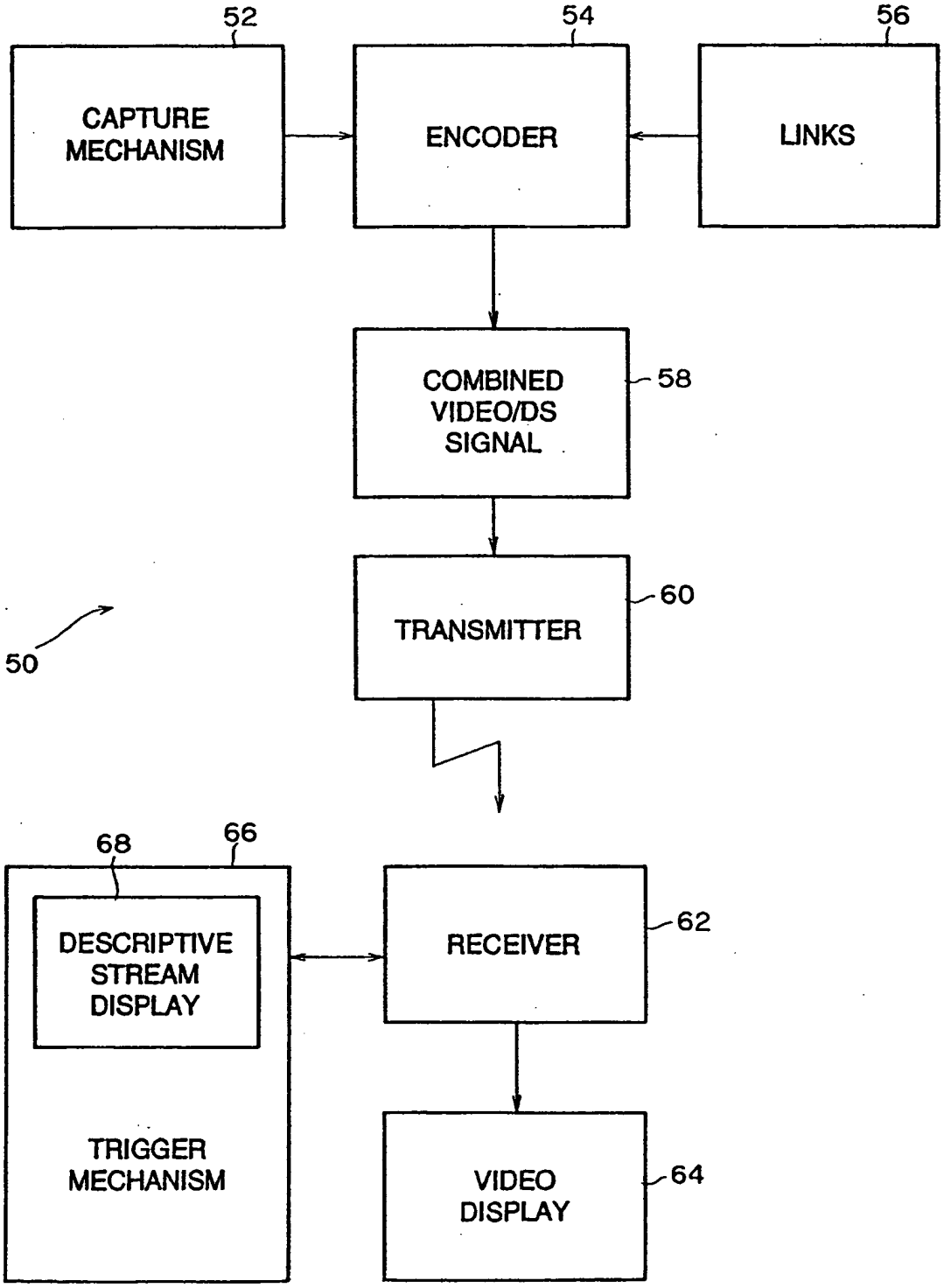

26
34
36
Object Index
Textual Description
Voice Annotation
Image Features
Object Links
URL Links
Java Applets
http://www.xyz.com/
38
40
Company Homepage
Product Guide
Awards
More Popcorn
etc..
28
Object Index
Textual Description
Voice Annotation
Image Features
Object Links
URL Links
Java Applets
http://www.studio.com/~actress
30
32
Actress Homepage
Other movie appearance
Oscar Awards
Backstage
etc..

FIG.3


52
54
56
CAPTURE MECHANISM
ENCODER
LINKS
COMBINED VIDEO/DS SIGNAL
58
TRANSMITTER
60
50
68
66
DESCRIPTIVE STREAM DISPLAY
RECEIVER
62
TRIGGER MECHANISM
VIDEO DISPLAY
64

INTERNATIONAL SEARCH REPORT

I. ational Application No
PCT/JP 98/01736

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F17/30

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimumdocumentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 12342 A (WISTENDAHL DOUGLASS A ;CHONG LEIGHTON K (US)) 3 April 1997 see page 21, line 7 - page 22, line 23; claims 1-6; figures ---	1-18
A	EP 0 618 526 A (US WEST ADVANCED TECH INC) 5 October 1994 see column 4, line 5 - column 6, line 28; figure 3 see column 9, line 33 - column 12, line 5 ---	1-18
A	"MULTIMEDIA HYPERVIDEO LINKS FOR FULL MOTION VIDEOS" IBM TECHNICAL DISCLOSURE BULLETIN, vol. 37, no. 4A, 1 April 1994, page 95 XP000446196 see the whole document ---	1-18

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of theinternational search	Date of mailing of the international search report
17 August 1998	25/08/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Fournier, C

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

national Application No
PCT/JP 98/01736

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BURRILL V ET AL: "TIME-VARYING SENSITIVE REGIONS IN DYNAMIC MULTIMEDIA OBJECTS: A PRAGMATIC APPROACH TO CONTENT BASED RETRIEVAL FROM VIDEO" INFORMATION AND SOFTWARE TECHNOLOGY, vol. 36, no. 4, 1 January 1994, pages 213-223, XP000572844 see the whole document	1-18

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

national Application No
PCT/JP 98/01736

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9712342 A	03-04-1997	US 5708845 A	13-01-1998
EP 0618526 A	05-10-1994	US 5442456 A	15-08-1995